

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 1999, 20:01:53 ; Search time 167.03 Seconds

(without alignments)
6306.250 Million cell updates/sec

Title: US-09-049-695-3

Perfect score: 534
Sequence: 1 TCCCGGCGCTGCTGGGGTTTA.....TCCACAGAGAGAAAAACAC 534

Scoring table:

IDENTITY_NUC
2546578 seqs, 986266752 residues

Phase :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	533	99.8	534	34	AA502331	AA502331 ne26h06.s
2	410.6	76.9	501	34	AA503839	AA503839 ne26h06.s
3	385.4	72.2	474	35	AA568450	AA568450 nm25b08.s
4	385	72.1	389	41	A1017393	A1017393 ou92f09.x
5	278.2	52.1	457	21	T85589	T85589 yd82g10.s1
6	264	49.4	418	50	A1699382	A1699382 wa69e01.x
7	254	47.6	424	21	T78178	T78178 yd79c11.r1
8	246.4	46.1	351	21	T85588	T85588 yd82g10.r1
9	194.4	36.4	378	21	T72043	T72043 yc63b05.r1
10	184	34.5	188	32	AA335186	AA335186 EST39634
11	178.8	33.5	357	21	T86494	T86494 yd81h10.r1
12	169.2	31.7	274	31	AA299977	AA299977 EST12598
13	139	26.0	139	35	AA551860	AA551860 ng01c09.s
14	127.8	23.9	590	44	A1323091	A1323091 m73b06.y
15	126.2	23.6	560	45	A1391147	A1391147 mc16g01.y
16	124.2	23.3	557	27	AA030782	AA030782 m15c06.r
17	123.2	23.1	586	39	AA871263	AA871263 vq33b08.r
18	121.6	22.8	571	28	AA060119	AA060119 m73b06.r
19	120	22.5	538	26	W30342	W30342 mc25c10.r1
20	117.6	22.0	600	25	W10974	W10974 ma46c11.r1
21	115	21.5	578	26	W36494	W36494 mb71h10.r1
22	110.4	20.7	510	39	AA871980	AA871980 vq47h04.r
23	110.2	20.6	508	39	AA871942	AA871942 vq42h10.r
24	109.4	20.5	485	37	AA674877	AA674877 vq51d08.r
25	108.8	20.4	498	39	AA870378	AA870378 vq44f10.r
26	107.6	20.1	478	37	AA691263	AA691263 v132h09.r
27	104	19.5	572	39	AA871332	AA871332 vq34a08.r
28	101.4	19.0	1114	26	W29238	W29238 mb96e09.r1
29	100.8	18.9	463	26	W17574	W17574 mb74f08.r1
30	100.2	18.8	543	28	AA060392	AA060392 m568b05.r
31	98.2	18.4	456	39	AA871485	AA871485 vq36a07.r
32	89.2	16.7	565	26	W36411	W36411 mb83g11.r1
33	87.2	16.3	387	49	A1639531	A1639531 rx02618s
34	84.4	15.8	479	25	W10393	W10393 ma40e07.r1
35	79.2	14.8	403	37	AA688510	AA688510 vq53f07.r
36	77.2	14.5	418	27	AA020563	AA020563 mh61c09.r
37	77	14.4	492	46	A1415324	A1415324 mc16g01.x
38	76.4	14.3	379	25	W14887	W14887 mb29f05.r1
39	75.4	14.1	497	26	W17866	W17866 mb83h03.r1
40	75.4	14.1	431	27	W08257	W08257 mb44b07.r1
41	69.4	13.0	285	50	AV035987	AV035987 AV035987
42	67.8	12.7	373	39	AA871696	AA871696 vq39g09.r
43	63.8	11.9	301	50	AV037226	AV037226 AV037226
44	60.8	11.4	423	26	W20636	W20636 mc16g01.r1
45	60	11.2	310	50	AV034645	AV034645 AV034645

ALIGNMENTS

RESULT 1
LOCUS AA502331
DEFINITION ne26h06.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:888523 3', similar to TR:G899433 G899433 PROLINE-RICH ACIDIC PROTEIN. ;, mRNA
ACCESSION AA502331
NID g2237298

VERSION AA502331.1 GI:2237298
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 534)
 NCBI-GenBank http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1404984.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCBI-GenBank clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html

Insert Length: 690 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 398.

FEATURES
 source
 1..534
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:898523"
 /clone_lib="NCI_CGAP_Co3"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /note="Vector: pRT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pRT73 vector. Library went through one round of
 normalization."

BASE COUNT 80 a 160 c 178 g 115 t 1 others
 ORIGIN

Query Match 99.8%; Score 533; DB 34; Length 534;
 Best Local Similarity 100.0%; Pred. No. 1.8e-113;

Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCCGGCTGCTGGGCTTATTGTCTAGTGGGCGAGGTTAGGAGGCTCCAGTCCCA 60
 DB 1 TCCCGGCTGCTGGGCTTATTGTCTAGTGGGCGAGGTTAGGAGGCTCCAGTCCCA 60
 OY 61 ACAGCCTGGGCTTGGGACAGGCGGGGCGAGTATGAGCCCTGAGACCTTACTGGGGGT 120
 DB 61 ACAGCCTGGGCTTGGGACAGGCGGGGCGAGTATGAGCCCTGAGACCTTACTGGGGGT 120
 OY 121 GGTAGATGTCGTCCTGCTTCCCTCCGGTCCAGAGACACCTGTATTTGGCATCACCC 180
 DB 121 GGTAGATGTCGTCCTGCTTCCCTCCGGTCCAGAGACACCTGTATTTGGCATCACCC 180
 OY 181 ACAACGGGGCTCTCTCTGCTGCTCTCAGCGGAGGGGTGGTACAGGCTGTCA 240
 DB 181 ACAACGGGGCTCTCTCTGCTGCTCTCAGCGGAGGGGTGGTACAGGCTGTCA 240
 OY 241 GGTGGGCTGAGGACTCAGGACAGGCGCGAGGTGCTCCTGCTCAGGCTGTG 300
 DB 241 GGTGGGCTGAGGACTCAGGACAGGCGCGAGGTGCTCCTGCTCAGGCTGTG 300

OY 301 TGCCTGGAAGGATGGGGCCCTGCCCCTGACCTGTCGCTTCTCCGTGGTCAAGAGTT 360
 DB 301 TGCCTGGAAGGATGGGGCCCTGCCCCTGACCTGTCGCTTCTCCGTGGTCAAGAGTT 360
 OY 361 TGGGCTTGGACAGGGAACACACACACAGCTGGTGTCTTCTCCGAGGCTCCACCA 420
 DB 361 TGGGCTTGGACAGGGAACACACACACAGCTGGTGTCTTCTCCGAGGCTCCACCA 420
 OY 421 CACGGGCGCCCGAGCTTCTGCGGCTGCTGTCGTGAGNGCAGTGTGACTTGCATCTT 480
 DB 421 CACGGGCGCCCGAGCTTCTGCGGCTGCTGTCGTGAGNGCAGTGTGACTTGCATCTT 480
 OY 481 GATAGGGAACCTTGGTGTCTGGGAATGCACCTGCTCCACAGAGAAAAACAAC 534
 DB 481 GATAGGGAACCTTGGTGTCTGGGAATGCACCTGCTCCACAGAGAAAAACAAC 534

RESULT 2

AA503839 501 bp mRNA EST 19-AUG-1997
 LOCUS ne53a10.s1 NCI_CGAP_Co3 Homo sapiens CDNA clone IMAGE:901050 3'
 DEFINITION similar to TR:G899433 G899433 PROLINE-RICH ACIDIC PROTEIN. ;, mRNA
 sequence.

ACCESSION AA503839
 NID 92238806
 VERSION AA503839.1 GI:2238806
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 501)
 NCBI-GenBank http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Nov 29, 1993 this sequence version replaced gi:637564.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-GenBank clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html

Insert Length: 586 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 190.

FEATURES
 source
 1..501
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:901050"
 /clone_lib="NCI_CGAP_Co3"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"

/note="Vector: pRT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pRT73 vector. Library went through one round of
 normalization."

BASE COUNT 62 a 148 c 165 g 126 t
 ORIGIN

Query Match 76.9%; Score 410.6; DB 34; Length 501;
Best Local Similarity 92.4%; Pred. No. 2.5e-85;
Matches 475; Conservative 0; Mismatches 10; Indels 29; Gaps 3;

```
QY 1 TCCGCGCTCTCTGGGTTATTGTTCTAGCTGGGGGAGGGTAGGAGGCTCCAGTCCCA 60
    |||||||
Db 16 TCCGCGCTCTCTGGGTTATTGTTCTAGCCCGGGGAGGGTCCGAGTCCCA 75
    |||||||
QY 61 ACAGCCTGGGCTTGGGACAGGGGGGAGTGTATGCTGAGCCCTGAGCCCTACTGGGGGT 120
    |||||||
Db 76 ACAGCCTGGGCTTGGGACAGGGGGGAGTGTATG-CCCTGAGCCCTACTGGGGGT 134
    |||||||
QY 121 GGTAGATGTGTCTTGTCTCTCCGCTCCAGAGACACCTGTGTATTGGATCACCC 180
    |||||||
Db 135 GGTAGATGTGTCTTGTCTCTCCGCTCCAGAGACACCTGTGTATTGGATCACCC 194
    |||||||
QY 181 ACAACCGGGGCTCTCTCTCCGCTCTCTCAGGCGGAGGTTGTACAGGCTGTCTAT 240
    |||||||
    195 ACAACCGGGGCTCTCTCTCCGCTCTCTCAGGCGGAGGTTGTACAGGCTGTCTAT 254
    |||||||
    241 GGTGCGGCTGGGACTGAGACAGAGGCCAGGGTGTCTGGTCTCATCCAGGCTTGG 300
    |||||||
    255 GGTGCGGCTGGGACTGAGACAGAGGCCAGGGTGTCTGGTCTCATCCAGGCTTGG 314
    |||||||
QY 301 TGCCTGGAGAGATGGGGCCCTGACCCTGCTGCTTCTCTCGGTGATCAAGATT 360
    |||||||
Db 315 TG-----CCTGCTGCTTCTCTCGGTGATCAAGATT 347
    |||||||
QY 361 TCGGCTTCTGACAGGAGACAGCACACAGCTGTGTCTCTTCTCCGAGGCTCACCA 420
    |||||||
Db 348 TCGGCTTCTGACAGGAGACAGCACACAGCTGTGTCTCTTCTCCGAGGCTCACCA 407
    |||||||
QY 421 CACGGGGGCCCCAGG-CTTCTGTGGGTCTGCTGTGAGNGGAGTGTACTGTGATCT 479
    |||||||
Db 408 CACGGGGGCCCCAGGTTCTTCTGTGGGTCTGCTGTGAGNGGAGTGTACTGTGATCT 467
    |||||||
QY 480 TGATAGGGAATTGGGTGCTGGGAATGACACTTG 513
    |||||||
Db 468 TGATAGGGAATTGGGTGCTGGGAATGACACTTG 501
    |||||||
```

RESULT 3
AA568450 474 bp mRNA EST 09-SEP-1997
LOCUS nm25b08.s1 NCI-CGAP_Col0 Homo sapiens cDNA clone IMAGE:106115 3'
DEFINITION similar to TR:G899433 G899433 PROLINE-RICH ACIDIC PROTEIN.; mRNA
sequence.
AA568450
92341504
AA568450.1 GI:2341504
EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 474)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397615.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:

www.bio.lnl.gov/db/ftp/image/image.html
Insert Length: 651 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 197.
Location/Qualifiers

FEATURES

source

1. 474

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13q13"
/clone="IMAGE:1061175"
/clone_lib="NCI-CGAP_Col0"
/issue_type="Colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pUT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pUT3D
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo (N-Soares4)."
BASE COUNT 61 a 142 c 157 g 114 t
ORIGIN

Query Match 72.2%; Score 385.4; DB 35; Length 474;
Best Local Similarity 98.5%; Pred. No. 1.5e-79;
Matches 452; Conservative 0; Mismatches 1; Indels 6; Gaps 6;

```
QY 2 CCCGGCTGTGGGTTATTGTTCTAGCTGGGGGAGGAGGAGGTCCTCCAGTCCAA 61
    |||||||
Db 18 CCCGGCTGTGGGTTATTGTTCTAGCTGGGGGAGGAGGAGGAGG-CCCACTCCAA 76
    |||||||
QY 62 CACCTGTGGCTTGGGACAGGGGGGAGTGTATGATGATGATGATGATGATGATG 121
    |||||||
Db 77 CACCTGTGGCTTGGGACAGGGGGGAGTGTATGATGATGATGATGATGATGATG 135
    |||||||
QY 122 GTAGATGTGTGTGGTCTCTCTGCTGCTCCAGACACCTGTGTATTGGATCACCA 181
    |||||||
Db 136 GTAGATGTGTGTGGTCTCTCTGCTGCTCCAGACACCTGTGTATTGGATCACCA 195
    |||||||
QY 182 CAACCGGGGCTCTCTCGGCTGTGCTCTCTCAGCGGAGGTTGTATGAGGCTGTATG 241
    |||||||
Db 196 CAACCGGGGCTCTCTCGGCTGTGCTCTCTCAGCGGAGGTTGTATGAGGCTGTATG 255
    |||||||
QY 242 GTGGGCTGTGGGACTAGAGACAGGCCCAAGGTGTCTGTGCTCATCAGGCTTGGT 301
    |||||||
Db 256 GTGGGCTGTGGGACTAGAGACAGGCCCAAGGTGTCTGTGCTCATCAGGCTTGGT 315
    |||||||
QY 302 GCCTGGAAGG-ATGGGGGCCCCGCTGACCTC-GTGGCTTCTCTC-GTGGTCAAGAG 358
    |||||||
Db 316 GCCTGGAAGGATGGGGGCCCCGCTGACCTCCTGTGGCTTCTCTCAGGCTCAAGAG 375
    |||||||
QY 359 TTTCGGCTTGTGACAGGAGAACACACACAGCTGTGTCTTCTCCGAGGCTCCAC 418
    |||||||
Db 376 TTTCGGCTTGTGACAGGAGAACACACACAGCTGTGTCTTCTCCGAGGCTCCAC 435
    |||||||
QY 419 CACACGGGGGCCCCAGG-CTTCTGTGGGTCTGTCTGA 456
    |||||||
Db 436 CACACGGGGGCCCCAGGCTTCTGTGGGTCTGTCTGA 474
    |||||||
```

RESULT 4
A1017393 389 bp mRNA EST 27-AUG-1998
LOCUS ou92f09.x1 NCI-CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1635305 3'
DEFINITION similar to TR:Q60874 Q60874 PROLINE-RICH ACIDIC PROTEIN.; mRNA
sequence.
A1017393
93231729
A1017393.1 GI:3231729
ACCESSION
NID
VERSION
KEYWORDS EST.

Db 61 ACAGCCTGGGCTTGGACAGAGCGGGGCGACGTATGCCCCCTTGAGAGCCCTACTGCGGGGT 120
QY 121 GGATAGATGTGCTTGTCTTCTCTCCGATCCGAGAGACCTGATGATTTGGATCAACC 180
Db 121 GGATAGATGTGCTTGTCTTCTCTCCGATCCGAGAGACCTGATGATTTGGATCAACC 180
QY 181 ACAACCGGGGCTTCTCTCGCCCT-6GTCTCTCTCAGCGGAGGATGTGTACAGGCTGTCA 239
Db 181 ACAACCGGGGCTTCTCTCGCCCTGCGCTCCTCAGCGGAGGATGTGTACAGGCTGTCA 240
QY 240 TGGTCGGGCTTGGAGCTCAGACAGCGGCCGAGGATGCTCGCTCATCAGAGCCCTTG 299
Db 241 TGGTCGGGCTTGGAGCTCAGACAGCGGCCGAGGATGCTCGCTCATCAGAGCCCTTG 300
QY 300 GTGCGCTGGAAGG--ATGCGGCGCCCTGCGCTGACCTGCTGAGCTTC-TCCTGCTGTCAC 355
Db 301 GGTGCTGTGAAGAAGATGGGGGCCCCCTGCTGACCTGCTGCTTCTTCTGCTGCTGTCAC 360
QY 356 GAGTTTGGGCTT 367
361 AANTTTCGGCTT 372

RESULT 6
LOCUS A1699382 418 bp mRNA EST 03-JUN-1999
DEFINITION wa69601.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
IMAGE:2301432.3 similar to contains Alu repetitive
element:contains element PTR5 repetitive element ;, mRNA sequence.
ACCESSION A1699382
NID 94987282
VERSION A1699382.1 GI:4987282
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 418)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135445.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from G1bco
High quality sequence stop: 368.
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2301432"
/clone_1lb="Soares_NFL.T.GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1 M.A.G.E. clones 297480-302087, 682632-687239,
725408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

61 a 113 c 141 g 103 t

Query Match 49.4%; Score 264; DB 50; Length 418;
Best Local Similarity 99.6%; Pred. No. 1.3e-51;
Matches 275; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 2 CCGGCGCTGCTGGGTTTATTGTCTAGCTGGGCGAGGGATGAGGAGGTCCTCCACCA 61
Db 144 CCGGCGCTGCTGGGTTTATTGTCTAGCTGGGCGAGGGATGAGGAGGTCCTCCACCA 203
QY 62 CAGCCTGGGCTTGGGACAGGGGGGCGAGTATGCCCCCTGAGCCCTACTGCGGGGTG 121
Db 204 CAGCCTGGGCTTGGGACAGGGGGGCGAGTATGCCCCCTGAGCCCTACTGCGGGGTG 263
QY 122 GTAGATGTGCTTGTGCTTCTCTCCGATCCGAGAGACCTGATGATTTGGATCAACCA 181
Db 264 GTAGATGTGCTTGTGCTTCTCTCCGATCCGAGAGACCTGATGATTTGGATCAACCA 323
QY 182 CAACCGGGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241
Db 324 CAACCGGGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382
QY 242 GTGCGGCTCGGAGCTCAGACAGCGGCCAGGCTGTC 277
Db 383 GTGCGGCTCGGAGCTCAGACAGCGGCCAGGCTGTC 418

RESULT 7
LOCUS T78178/c 424 bp mRNA EST 15-MAR-1995
DEFINITION yd79c11.r1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone
IMAGE:114452.5 similar to contains PIR5 repetitive element ;, mRNA
sequence.
ACCESSION T78178
NID 9696687
VERSION T78178.1 GI:696687
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 424)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL
COMMENT
TITLE
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 644
High quality sequence stops: 403 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 644 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 403.
Location/Qualifiers
1..424
/organism="Homo sapiens"
/db_xref="GDB:470069"
/db_xref="taxon:9606"
/clone="IMAGE:114452"
/clone_1lb="Soares fetal liver spleen INFUS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and spleen; Vector: pT7T3D (Pharmacia)"

FEATURES
source

JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
High quality sequence stops: 291
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: M13RP1
High quality sequence stop: 291.
Location/Qualifiers
1. 378
/organism="Homo sapiens"
/db_xref="GDB:502386"
/db_xref="taxon:9606"
/clone="IMAGE:85329"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Hepatectomy from normal male caucasian. Average
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 83 a 117 c 116 g 59 t 3 others

ORIGIN

Query Match 36.4%; Score 194.4; DB 21; Length 378;
Best Local Similarity 82.8%; Pred. No. 1.3e-35;
Matches 256; Conservative 0; Mismatches 24; Indels 29; Gaps 2;

QY 229 ACAGGCTGTATGTCGGGCTCGGAGTCAAGACAC---GGCCAGGGTGTCTCGGCT 285
DB 373 ACAGGCTGTATGTCGGGCTCGGAGTCAAGACACGGCCAGGGTGTCTCGGCTCT 314
QY 286 CCATCCAGGCTTGTGCTCGGAGATGGGGCCCTGACCTGCTGCTTCTCT 345
DB 313 CCATCCAGGCTTGTGCTCGGAGATGGGGCCCTGACCTGCTGCTTCTCT 280
QY 346 CGGTGCTCAAGAGTTTGGCTTCTGACAGGGAACACACCAGCTGCTGCTTCT 405
DB 279 CGGTGCTCAAGAGTTTGGCTTCTGACAGGGAACACACCAGCTGCTGCTTCT 220
QY 406 CCGAGGCTCCACACACAGGGGCGCCAGGCTTCTGCGGCTCGCTGAGNGCAGTG 465
DB 219 CCGAGGCTCCACACACAGGGGCGCCAGGCTTCTGCGGCTCGCTGAGNGCAGTG 160
QY 466 TTGACTTGCATCTTGATAGGAAGTGGTGTCTGGAGATGCACCTGCTCCACAGAAG 525
DB 159 TTGACTTGCATCTTGATAGGAAGTGGTGTCTGGAGATGCACCTGCTCCACAGAAG 100
QY 526 AAAAACAC 534
DB 99 CACAACNAC 91

RESULT 10
AA335186/c 188 bp mRNA EST 21-APR-1997
LOCUS AA335186
DEFINITION EST39634 Epididymus Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA335186
NID 91987501
VERSION AA335186.1 GI:1987501
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 188)
AUTHORS
Adams,M.D., Kellavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marnaros,S.M., Merrick,M.,
Moreno-Balaguera,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Hasetline,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL 96026280
MEDLINE
COMMENT
On Nov 29, 1993 this sequence version replaced gi:430523.
Other ESTs: TH0104700
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1..188
/organism="Homo sapiens"
/db_xref="ATCC (inhost):136917"
/db_xref="taxon:9606"
/map="21"
/clone_lib="Epididymus"
/sex="male"
/dev_stage="adult"
/note="Organ: Epididymis; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 40 a 68 c 51 g 26 t 3 others

ORIGIN

Query Match 34.5%; Score 184; DB 32; Length 188;
Best Local Similarity 98.4%; Pred. No. 2.5e-33;
Matches 184; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGGGCTGTGGGGTTTATTGTCTAGCTGGGGCAGGGTAGGGAGGGTCCCACTCCAA 61
DB 188 CCGGGCTGTGGGGTTTATTGTCTAGCTGGGGCAGGGTAGGGAGGGTCCCACTCCAA 129
QY 62 CAGCTGGGCTTGGGACAGAGGGGGGAGTATGATGCCCCCTGAGCCCTACTAGGGGGTG 121
DB 128 CAGCTGGGCTTGGGACAGAGGGGGGAGTATGATGCCCCCTGAGCCCTACTAGGGGGTG 69
QY 122 GTAGATGTGCTTGTGCTCTCCGTCAGGACAGACCTGTGATTTGGCATCACCA 181
DB 68 GTAGATGTGCTTGTGCTCTCCGTCAGGACAGACCTGTGATTTGGCATCACCA 9
QY 182 CAACCG 188
DB 8 MAACCG 2

Matches 185: Conservative 0: Mismatches 12: Indels 1: Gaps 1:

QY 338 CTTCCTCGGTGTCACAGAGTTTCGGCTTCTGACAGGGAACAGCACCAGCTGATC 397
 |||||||
 Db 274 CTTCCTCGGTGTCACAAAGTTTCGGCTTCTGACAGGGAACAGCACCAGCTGATC 215
 |||||||
 QY 398 GTCTCTTCGCGAGGCTCCACACAGGCGGCCCGCAGG-CTTCTCTGGTCTCTGCTGA 456
 |||||||
 Db 214 GTCTCTTCGCGAGGCTCCACCAAGCGGCGGCCCGCAGGCTCTCTGGTCTCTGA 155
 |||||||
 QY 457 GGGGAGTGTTCAGTCTGTGATAGGAGACTGGGTGGGGAATGACACCTGGCTC 516
 |||||||
 Db 154 GGGCAGAGTGTTCAGTCTGTGATAGGAGACTGGGTGGGAGCTGACCTGCTCC 95
 |||||||
 QY 517 CCACAGAGAAAAACAC 534
 |||||||
 Db 94 CCACAGAGACAGNACACAC 77
 |||||||

RESULT 13
 1860

AA51860 139 bp mRNA EST 04-SEP-1997
 DEFINITION nc01009.s1 NCI CGAP Co3 Homo sapiens CDNA clone IMAGE:928144.3'
 similar to TR:G899433 G899433 PROLINE-RICH ACIDIC PROTEIN. ;, mRNA
 sequence.
 AA51860
 MID 92322112
 VERSION AA51860.1 GI:2122112
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 139)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Apr 14, 1993 this sequence version replaced gi:315285.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbrrp/image/image.html

Trace considered overall poor quality
 Insert Length: 547 Std Error: 0.00
 Seq primer: -40m13 fwd. Et from Amersham
 High quality sequence stop: 1.

FEATURES

source
 1..139
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:928144"
 /clone_lib="NCI_CGAP_Co3"
 /sex="pooled"
 /tissue.type="colon"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded CDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library went through one round of
 normalization."

BASE COUNT 17 a 42 c 47 g 33 t
 ORIGIN

Query Match 26.0%; Score 139; DB 35; Length 139;
 Best Local Similarity 100.0%; Pred. No. 5.1e-23;
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 GGTGTAGATGTGATGCTGTGCTTCTCTCGGTCACAGAGACACGTGATTTGGCATCA 177
 |||||||
 Db 1 GGTGTAGATGTGATGCTGTGCTTCTCTCGGTCACAGAGACACCTGATTTGGCATCA 60
 |||||||
 QY 178 CCCACACCGGGGCGCTCTCTGCTCCCTGATCTCTCAAGCGGAGGTGTACAGCTGT 237
 |||||||
 Db 61 CCCACACCGGGGCGCTCTCTGCTCCCTGATCTCTCAAGCGGAGGTGTACAGCTGT 120
 |||||||
 QY 238 CATGGTCGGGCTCGGGA 256
 |||||||
 Db 121 CATGGTCGGGCTCGGGA 139
 |||||||

RESULT 14

A1323091 590 bp mRNA EST 23-DEC-1998
 LOCUS m173b06.y1 Soares mouse p3NMF19.5 Mus musculus CDNA clone
 DEFINITION IMAGE:481715 5' similar to gb:U28486 Mus musculus uterine-specific
 prolone-rich acidic protein (MOUSE);, mRNA sequence.
 A1323091
 MID 94057520
 VERSION A1323091.1 GI:4057520
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 590)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Apr 14, 1993 this sequence version replaced gi:639224.

Contact: Marra Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:292459

This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -40RP from Glibco
 High quality sequence stop: 444.

FEATURES

source
 1..590
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:481715"
 /clone_lib="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT7T3D (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTATCCAAATCTGAAGTGGAGCGCGCAATTTTATTTTATTTT 3'],
 double-stranded CDNA was size selected, ligated to Eco RI

Db	146	CCTTGGTCTTGACAGGAGCTGGTGTGCTGGAGCTGGCCTGGCTCCACACAGAGTGA	87
Oy	530	ACAAC	534
Db	86	GGCAC	82

Search completed: August 19, 1999, 20:45:34
Job time: 2621 sec
